GATR Journal of Business and Economics Review

GATR JOURNALS

Journal homepage: http://gatrenterprise.com/GATRJournals/JBER/vol7_2022_issue1.html

J. Bus. Econ. Review 7(1) 61 – 74 (2022)

Predicting Trade Mispricing: A Gaussian Multivariate Anomaly Detection Model

Olalere Isaac Opeyemi[1*], Dewa Mendon[2], Dlamini Lenhle[3]

[1,2,3]Durban University of Tech., Durban, 4000, Durban, KZN, South Africa

ABSTRACT

Objective - This paper predicts a measurement indicator for the trade mispricing channel and its effectiveness in identifying IFFs.
Methodology – A model gaussian multivariate anomaly detection algorithm, for classifying between legal and illegal transactions that are suspicious in terms of misreporting was developed. The method is a machine learning technique and uses data from South Africa, Botswana, the USA, and China over a period from 2000 to 2019, to learn whether there are any intriguing differences in the model performance based on these countries and the effect of other factors. Imports and Exports are used as features of the model while the net flow derived from these features is used as the third feature of the model. Imports and exports data are sourced from IMF's Direction of Trade Statistics database. Annual tariffs data and corruption data come from the WDI database and Transparency International's Corruption Perception Index, respectively. Data for 'accounting and auditing standards' comes from the world economic forum.
Findings - The result showed that while the model may be effective in detecting IFFs due to mispricing, other factors may however contribute to irregularities of trading data that is flagged as IFFs. This in addition to accounting for total quantum, also provides details empowering governments with the information to stimulate and drive the desire to curb IFFs from its different sources and channels.
Novelty - This study contributes to the debate on trade mispricing by proving a baseline measurement to help detect and track IFFs.
Type of Paper: Empirical
JEL Classification: F17, Q02
Keywords: Gaussian Multivariate Anomaly Detection; GMAD; Illicit Financial Flow; IFF., Trade Mispricing;

Reference to this paper should be made as follows: Opeyemi, O.I; Mendon, D; Lenhle, D. (2022). Predicting Trade Mispricing: A Gaussian Multivariate Anomaly Detection Model, *J. Bus. Econ. Review*, 7(1), 61–74. https://doi.org/10.35609/jber.2022.7.1(2)

1. Introduction

Africa is estimated to have lost over 1 trillion US dollars in Illicit financial flows (IFFs) for the past five decades (Kar & LeBlanc, 2013). This loss has hindered both economic and human development. Even though the actual size of the illicit financial flows is unknown and uncertain, potentially due to poor record practices and quality data in African Countries. Illicit financial flow data is grouped into three components, which are criminal activities, commercial activities, and corruption. According to Africa (2015), commercial activities account for 65% of illicit financial flows in Africa.

[*] Paper Info: Revised: April 16, 2022
Accepted: June 30, 2022
[*] Corresponding author: Olalere Isaac Opeyemi
E-mail: isaaco@dut.ac.za
Affiliation: Durban University of Tech., Durban, 4000, Durban, KZN, South Africa

Commercial activities include tax avoidance, trade mispricing, aggressive transfer pricing and tax evasion (Cobham & Janský, 2018). The GFI reports that approximately 60% of IFFs are channeled through trade mispricing (Kar, Cartwright-Smith, & Hollingshead, 2010). However, the value in 2015 rose to 83% (Kar & Freitas, 2012). Therefore, trade mispricing channels account for a large proportion of the illicit financial flows, it is a form of tax fraud including importers and exporters intentionally misreporting the quantity, value and the type of good or service in a commercial transaction. In literature, it is commonly referred to as Trade-Based Money Laundering (TBML).

TBML has been examined based on an assumption because of the absence of actual data on illicit financial flows. Therefore, an index is constructed that is not based on actual data but rather on the likelihood that the flows exceed some arbitrary threshold (Forstater, 2015). Some of the indices include Financial Secrecy Index and the Basel Anti-Money-Laundering Index. The drawback of these indices is that they cannot be tested against real data and might have an unprovable amount of errors (Forstater, 2017). As a result, some illicit financial flow classifications from different channels, for example, Trade-based money laundering might be detected by using a gaussian multivariate anomaly detection algorithm. This algorithm is not dependent on government compiled data, it is economical, mitigates missing data and resists quantification challenges. To the best of our knowledge, this is the first study to do so.

There is a rapidly expanding body of literature on the determinants of trade mispricing. One strand of studies focuses on how measures of the trade gap change the incentive to evade tax or tariffs (Fisman & Wei, 2004).Fisman and Wei (2004) found that import taxes on Hong Kong shipments to China is positively correlated with China's tax rates. There is also a negative correlation between tax rates of closely related products, implying that evasion takes place through misclassification. These results are corroborated in a transnational study by Kellenberg and Levinson (2019) and Bouet and Roy (2012) who focus on Nigeria, Kenya and Mauritius. In India, the enforcement of product characteristics that determine the ease of detection of evasion can limit the responsiveness of evasion to tariffs (Mishra, Subramanian, & Topalova, 2008). A similar study was conducted in the Philippines by Yang (2008) who found that enforcement reduced evasion gaps. However, these efforts displaced alternative duty avoidance methods. For instance, when the products are differentiated, the responsiveness of the trade gap to the tariff level is more than that of similar goods (Javorcik & Narciso, 2008). Furthermore, tariff evasion results from misrepresentation of import prices instead of product misclassifications. Rotunno and Vézina (2012) found that trade misinvoicing in China is prone to firms that are less likely to be scrutinized by that state. Similar to Tunisian firms that are politically connected, under-invoicing is prevalent (Rijkers, Baghdadi, & Raballand, 2017). Therefore, corruption is also a predictor of evasion gaps (Worku, Mendoza, & Wielhouwer, 2016). For instance, Sequeira (2016) found tariff rates, bride payments to customs officials and evasion gaps in Mozambique to South Africa trade corridor. Trade misinvoicing in exporting commodities can also result from export controls (Vézina, 2015). In order to curb evasion gaps, the World Trade Organization imposed custom valuation agreements (Javorcik & Narciso, 2017).

The majority of the literature examines the relationship between an estimate of IFFs and regressing it on a series of observable characteristics that might be correlated to IFFs. The interpretations rely on important identification assumptions. We argue that detecting IFF based on this kind of framework is likely to miss some of the key channels and sources within and across a border of countries. Therefore, our analysis considers a machine learning approach using an anomaly detection framework that allows us to detect anomalies by identifying transactions that deviate from what is considered normal.

The rest of the paper is organized as follows. Section 2 presents the literature review. Section 3 discusses methodology while section 4 outlines the results and discussions. Finally, section 5 is the conclusion of the paper.

2. Literature Review

Combating illicit financial flows has become an imperative agenda globally over the years. Consequently, the United Nations included the fight against illicit financial flows as one of the Sustainable Development Goals. According to Doumbia and Lauridsen (2019) illicit financial flows produce unequal opportunities both domestically by impeding the likelihood of development of poor countries and internationally by increasing wealth disparities. Tackling and measuring illicit financial flows is very difficult due to the nature of illicit financial flows not being systematically recorded. The challenges of quantification significantly hinder the understanding of the severity of the IFFs in particular trade misinvoicing. The available trade misinvoicing estimates have been scrutinized immensely for their robustness of methodology in policy recommendations against IFFs see for example Cobham and Janský (2018); (Nitsch, 2016).

Ample literature Fisman and Wei (2004); Kellenberg and Levinson (2019) have used methodologies that are limited due to large estimates of trade misinvoicing and are generally regarded as weak evidence of cross-border financial flows. Other literature Davies, Martin, Parenti, and Toubal (2018); (Wier, 2020) have used transaction-level international trade that is normally limited due to administrative restrictions on access to transaction-level customs data. However, other channels of IFFs have been obtained from aggregate economic data for example bilateral trade statistics that measure trade misinvoicing. For this reason, the use of machine learning has demonstrated significant reliability in detecting IFFs and abnormal patterns in statistics see for example (Wang & Han, 2019; Wohl & Kennedy, 2018; Wu, 1997). The main conclusion from these studies is that machine learning is able to detect IFFs and have predictive accuracy. For instance, Wohl and Kennedy (2018) show that the IFF cross-border (tax evasion) trade patterns predictive capacity is less superior when applying gravity models compared to machine learning. These findings are consistent with those Wu (1997) who found 94 percent accuracy in detecting potential audit violations using machine learning. This paper is related to Wang and Han (2019) who used a support vector machine (SVM) model to predict credit card fraud. However, this study focuses on trade misinvoicing and instead of employing an SVM we use anomaly detection which performs better. To develop, describe and validate a machine learning model for prioritizing which transaction should be investigated for potential IFFs.

2.1 Research problem

Illicit financial flows (IFFs) are a major problem for African developing countries, by draining their tax revenues and domestic capital. Therefore, suppressing their ability to mobilize scarce resources is required for improving the three dimensions of sustainable development (social, economic and environmental) and socio-economic growth. Illicit financial flows are driven by trade openness in the context of inadequate regulation and poor governance; poor regulation of the financial system and capital account; and inadequate institutional quality and excessive dependence on commodity exports (Ndikumana, 2014). These drivers serve as channels for the environmental, economic, institutional and social harm caused by illicit financial flows. The various channels are illegal and by nature hidden, clandestine activities. This poses enormous challenges in understanding the breadth of channels and mechanisms through which illicit financial flows are conducted and how they affect sustainable development in Africa. This study thus aims to classify and predict the channels through which illicit financial flows.

Channels of illicit financial flows can be classified from either their origin (source) or if their eventual use is illegal. Taking into account whether the channel is legal, there are different types of international financial flows classified as illicit financial flows. According to Baker (2005) illicit financial flows can be sourced from criminal proceeds which include individual tax evasion and drug money, corrupt proceeds, which include money stolen from the state and commercial proceeds, which include tariff evasion. All the sources of illicit financial flows are hidden, and they take many forms using different types of channels. Moreover, their measurement is challenging both conceptually and in practice. These challenges are country-specific and influenced by statistical practices, institutions and national priorities.

3. Methodology

The study uses a monthly data period from 2000 to 2019. Imports and Exports are used as features of the model while the net flow derived from these features is used as the third feature of the model. Imports and exports data are sourced from IMF's Direction of Trade Statistics database. Annual tariffs data and corruption data come from the WDI database and Transparency International's Corruption Perception Index, respectively. Data for 'accounting and auditing standards' comes from the world economic forum.

An anomaly detection algorithm identifies a class and checks whether a new feature falls within the normal classification based on existing trained data or is classified as an anomaly that requires further testing. Adopting a machine-learning artificial intelligence, (AI), approach for analysis of existing data may give more headway in detecting illicit financial flows and also identifying the most characteristics across different regions that effectively flag the flow. As compared to other novel methods, support vector machine, (SVM), used in literature, Anomaly detection performs better in that, it is a better algorithm when there are unequal samples of positive and negative examples.

Anomalous algorithm with gaussian distribution performs density estimation (DE) on a training set, $x^{(1)}, x^{(2)}, \dots, x^{(m)}$, and checks whether a new example, x_{test}, is abnormal/anomalous/illicit. This is achieved by developing a model (x) , that tells the probability that an example is anomalous/illicit. Gaussian distribution is parameterized by a mean, μ, and variance, σ^2 of the trained data. Hence, Equation 1, therefore, defines the mathematical expression for anomaly detection with a gaussian distribution.

$$P(X:\mu,\sigma^2)=\frac{1}{\sqrt{2\pi\sigma}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{1}$$

Given that mean, μ, is expressed as indicated in Equation 2 while variance is expressed in Equation 3.

$$\mu=\frac{1}{m}\sum_{i=1}^{m}x^{(i)} \tag{2}$$

$$\sigma^2=\frac{1}{m}\sum_{i=1}^{m}(x^{(i)}-\mu)^2 \tag{3}$$

There is a threshold called, epsilon (ϵ), which is the dividing line between normal examples and anomalous/illicit. A given example is anomalous when $P(x) < \epsilon$ and it is normal when $P(x) > \epsilon$. This implies that an example is flagged as a positive example (illicit financial flow) when $y = 1$, i.e $P(x) < \epsilon$, which is anomalous and as a negative example when $y = 0$, i.e, $P(x) > \epsilon$ which is a normal data point condition. The threshold ϵ, will most likely vary from one trained data set to another.

An anomaly detection algorithm is aimed at flagging Illicit Financial flows due to misreporting with factors such as tariffs, accounting and auditing standards, and corruption as possible indicators. The data used for the analysis are dependent on the models employed to derive the variables considered by the training algorithm. Firstly, net flows for the TBML model are used for computing the difference between Imports (Inflows) and Exports (Outflows) data.

$$Net\ flows\ for\ TBML = Outflows - Inflows \tag{4}$$

Secondly, another parameter considered for this analysis is the trade reporting gap which is represented in Equation 5.

$$Trade\ reporting\ gap = \frac{V_{xm}^{m} - V_{xm}^{x}}{V_{xm}^{m} + V_{xm}^{x}} \tag{5}$$

Where V_{xm}^{m}, is the quarterly total trade shipped from exporting country x to importing country m.

V_{xm}^{x}, is that same quarterly value as reported by the exporter x.

These two features are considered for model P(x) analysis for flagging whether there was illicit financial flow. This may also be referred to as an anomaly by the model.

Furthermore, the illicit financial flow may be triggered or detected by considering the factors, such as tariffs, accounting and auditing standards and corruption, which are a measure of a country's characteristics for the flagging anomaly. Equation 6 illustrates the 'log trade gap' variable as a function of some factors.

$$InV_{xm}^{m} - InV_{xm}^{x} = \beta_0 + \beta_1^1 \sigma_{xmt} + \beta_2 Z_{xmt}^{m} - \beta_3^1 Z_{xmt} + \mu_{xmt} \tag{6}$$

Where σ is the trade cost, μ is the mean-zero random error, $\beta_2$and β_3 are coefficient which should be zero if the difference, $\mathrm{InV}_{xm}^{m} - \mathrm{InV}_{xm}^{x}$, is not a function of tariffs, accounting standards, or corruption, i.e should differ only by CIF trade cost. Z is a measure of a country's characteristics. Log trade gap is also a feature that can be considered in evaluating whether or not a trade transaction is illicit using the Gaussian Multivariate Anomaly Detection (GMAD) model.

The corruption index of the considered countries may also indicate a very pertinent correlation in the data variation for import and export flows. While the model can be considered to independently analyze the financial flows based on the import and export data between two or more trading countries, it is however also important to analyze the variance between the model parameter that flags illicit financial flows in countries with high corruption index alongside a country with low corruption index. However, due to the unavailability of data for most African countries, two (2) African countries were considered for analysis, namely South Africa and Botswana. Another factor that was considered in the selected countries in the analysis is the trade relationship between the countries and the contrasting corruption index. Countries considered based on trade relationships and contrasting corruption indexes include the United States of America, Germany, and China. Trade flows between these partner countries and also between the African trading nations may indicate if there are any intriguing differences and effects of the factors based on these countries.

The data used for this research work are sourced from an existing data bank online and based on availability, quarterly and annual data are collected, over a period of 20 years. Imports and exports data are sourced from IMF's Direction of Trade Statistics database, which is used for calculating 'Net flows for TBML model' and 'trade reporting gap' in equations 4a and 4b respectively. Annual tariffs' data is gotten from the World Development Indicators (WDI), while the corruption data comes from Transparency International's Corruption Perception Index which scores countries on a scale of 0 (very corrupt) to 100 (not corrupt).

A high-level programming language, MATLAB/OCTAVE, is used for the analysis and the graphs showing the contours of the boundary of the normal and illicit flows are illustrated.

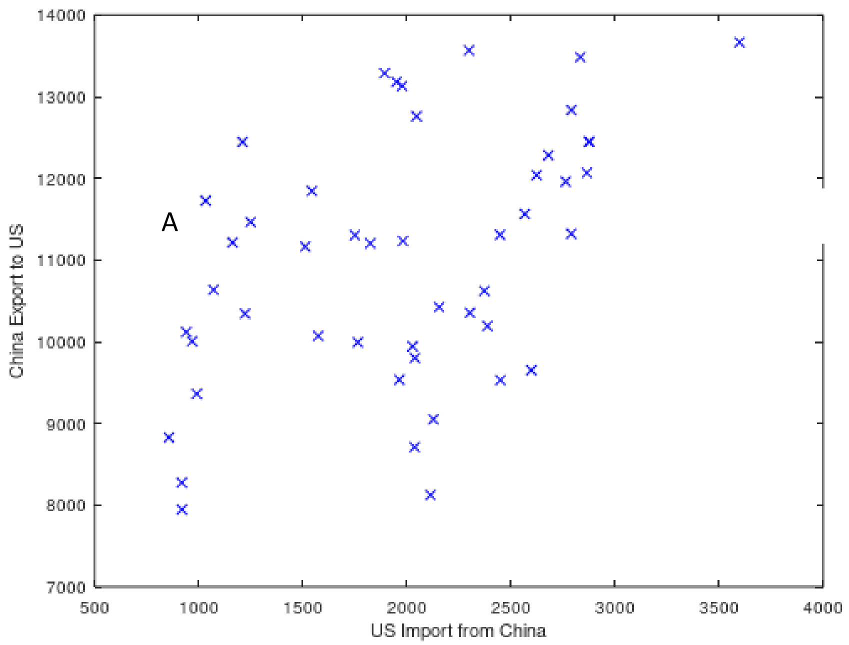
From equation 4, the netflows from the TBML model should account for misinvoicing of international trade flows, K_{ij}, which accounts for partner country estimates of exports. This is also referred to as CIF/FOB ratio. Where CIF is the cost, insurance and Freight factor while the FOB is free on-Board goods. This ratio may either be taken as 1.05 or 1.1 depending on the trading distance and destinations. It is however assumed that over or under-invoicing is considered to be trade mispricing.

4. Results and Discussion

4.1 Analysis Between China and the US Trade Flow

The trade flow between China and the US should follow an almost similar export and import trade reporting, except for the trade reporting gap that should account for about 1.05% of the FOB goods. Trade exports from China to the US can be plotted against the recorded same trade import from China by the US as indicated in Figure 1A. The trade reporting gap between the US and China trade is taken as 1.1% of the FOB goods. This accounts for the trade distance and CIF cost incurred during trade transactions.

This factor should accommodate for the difference between the export trade reported by the exporting country, China, as compared to the exact data recorded by the importing country, the US. The quarterly data is analyzed using the GMAD model to determine the best epsilon on the cross-validation set. The mean and variance of the netflow of the export and import trade are evaluated and recorded as parameters for determining the expression for indicating illicit financial flow by the model. Each of the required parameters of the model is evaluated by creating a function (codes) which are called from the main method.



B

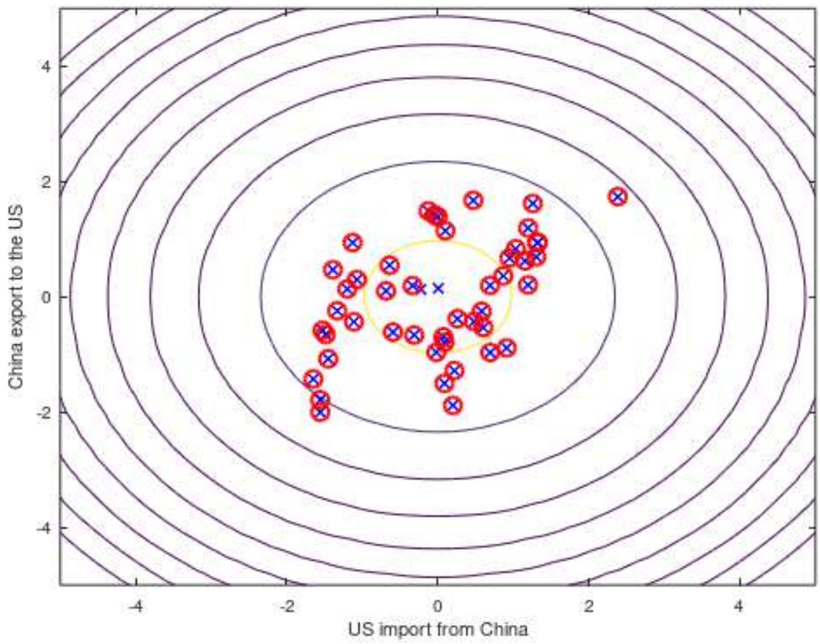


Figure 1: US- China Trade Flow

The mean, μ, of the netflow data between the United State quarterly import data from China and the corresponding recorded export of China to the US between the years 2000 and 2011 is 9075.3 while the evaluated variance is 1506166.69. The variation in the mean of the recorded import data of the US from China and the corresponding recorded export data of China is marginally noticeable. Equally, the variance of the netflow data also depicts a wide variance in the recorded data within the same time frame. As this may in a way indicate that illicit financial flow may be the result, it is also important to examine the model by analysing the parameters of the gaussian multivariate model. The epsilon parameter evaluated on the validation set data, which is the import data of the US from China from the year 2012 to the first quarter of 2018 and the corresponding recorded export of China to the US in the same period, is 0.4012623.

From equation 1, the values for both parameters can be substituted, and the expression becomes:

$$P(X:\mu,\sigma^2)=\frac{1}{\sqrt{2\pi*1506166.69}}\exp(-\frac{(x-9075.3)^2}{2*1506166.69})<\varepsilon(0.4012623) \qquad (7)$$

Where the variable 'x' in Equation 8 is the netflow data, which is the difference in each quarterly recorded US import from China and the corresponding quarterly recorded China export to the US. If the resulting value at any given quarterly data is less than the derived epsilon, then it is flagged as an illicit transaction.

The model evaluates illicit transactions on the data set and a graphical illustration of the trade transaction is shown in Figure 1B.

The graph indicated in Figure 1B shows that just two data points were not flagged as illicit financial flow using the model. This may also be correlated with the high variance between the mean and variance values of netflow data (between the US and China). There may be other factors and concerns that trigger the wide trade reporting gap between the two trading countries, but the model effectively flagged ninety-five percent of the trading data between the considered years as illicit. To further correlate the analytical method using the GMAD model, other trading countries are considered and analysed.

4.2 Analysis between South Africa and Botswana

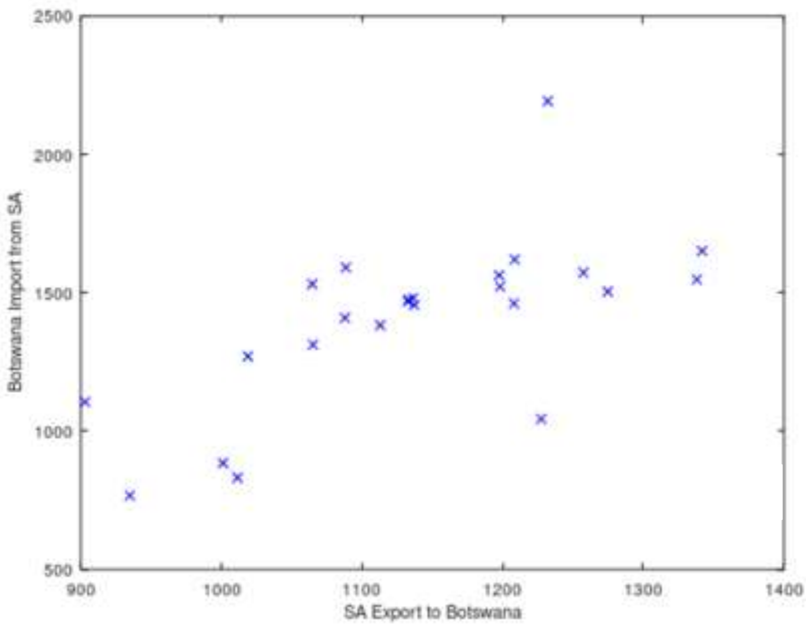
To further discuss the trade relationship between countries and their illicit financial flow, the trade relationship between two African countries is considered. This would evaluate the trend of financial flow between two developing economies and also show any similarity, if there is, in illicit financial flow pattern between two developing economies as compared with two developed economies of the world. The challenge however with most developing nations, especially in Africa, is the unavailability of data or a functional data capturing means, hence, making it a bit challenging to analyze and reconcile the result with other developed countries. However, there are quite a few African countries' trade (import and export) data available on IMF's Direction of Trade Statistics database. South Africa has quite a few trading partners among other African countries, but the data are mostly not available on IMF's Direction of Trade Statistics database.

This section discusses the analysis of the import and export data between South Africa and Botswana between the year 2000 and 2019 using the gaussian multivariate anomaly distribution AI method. As earlier analysed in previous sections, this section would only consider the quarterly import and export data between South Africa and Botswana between the year 2010 and 2019 due to the unavailability of the export data from South Africa to Botswana from 2000 to 2009. Figure 2A shows the graph of Botswana's quarterly imports from South Africa from 2010 to 2019 against the corresponding recorded South Africa's export to Botswana.

The trade reporting gap between South Africa and Botswana is taken as 1.05% for determining the trade label for the cross-validation set as illicit or not. The trade reporting gap value of 1.05% is used for the calculation because the distance between the trading or partner country is considerably small as compared to the trade across the continent. Similarly, the dataset is learnt and trained as earlier enumerated in the previous sections by dividing the dataset into training, validation and test set, and analysed using gaussian multivariate anomaly distribution for flagging illicit and normal trade transactions.

The model learned and trained on the training set using Gaussian distribution to perform the density estimation in order to determine the mean μ, of the data set and the variance σ^2. The evaluated mean, μ, is 53.307s and the variance, σ^2, of the model P(x), which tells the probability that an example is anomalous/illicit is 173822.69. The best epsilon found using cross-validation set 0.3868411 for flagging whether or not the condition for illicit financial

A



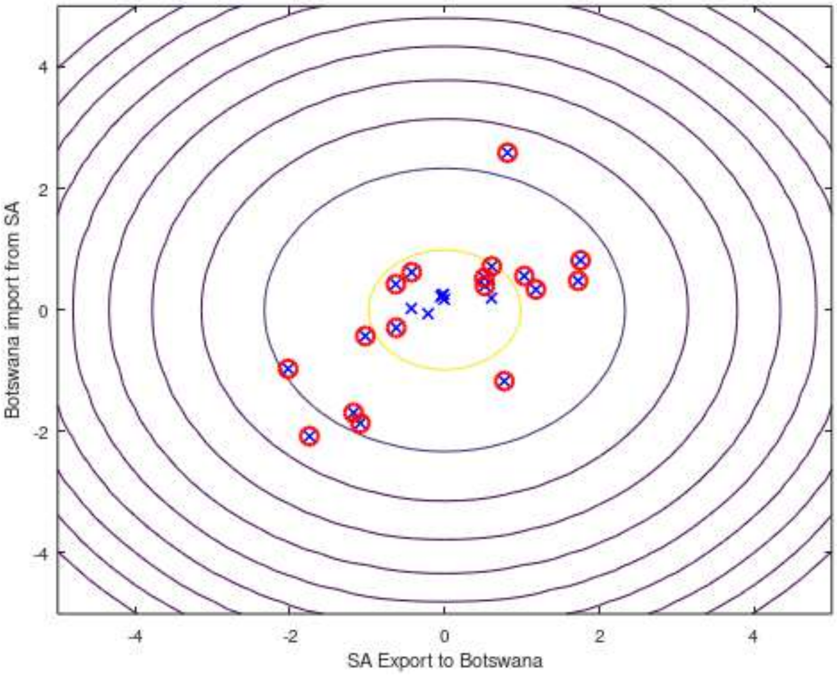


Figure 2: South Africa and Botswana Trade

Figure 2B presents the graph of the trade between South Africa and Botswana and the points marked as illicit transactions are indicated with red circles. With 73% of the data points flagged as illicit financial flows, this indicates that there may be some other factors, for example, tariff, policies, etc that extends the trade reporting gap beyond the normal acceptable gap or variance.

To further ascertain the capability of identifying illicit and non-illicit financial flows using the GMAD model, corresponding simulated data between South Africa and Botswana is considered. In this analysis, the simulated data of South Africa's export to Botswana is considered against the real data of Botswana's imports from South Africa. The simulated data is generated randomly considering the trade reporting gap of 1.05% less of the import data of Botswana from South Africa. The data is then plotted against the import of Botswana from South Africa as indicated in figure 3A.

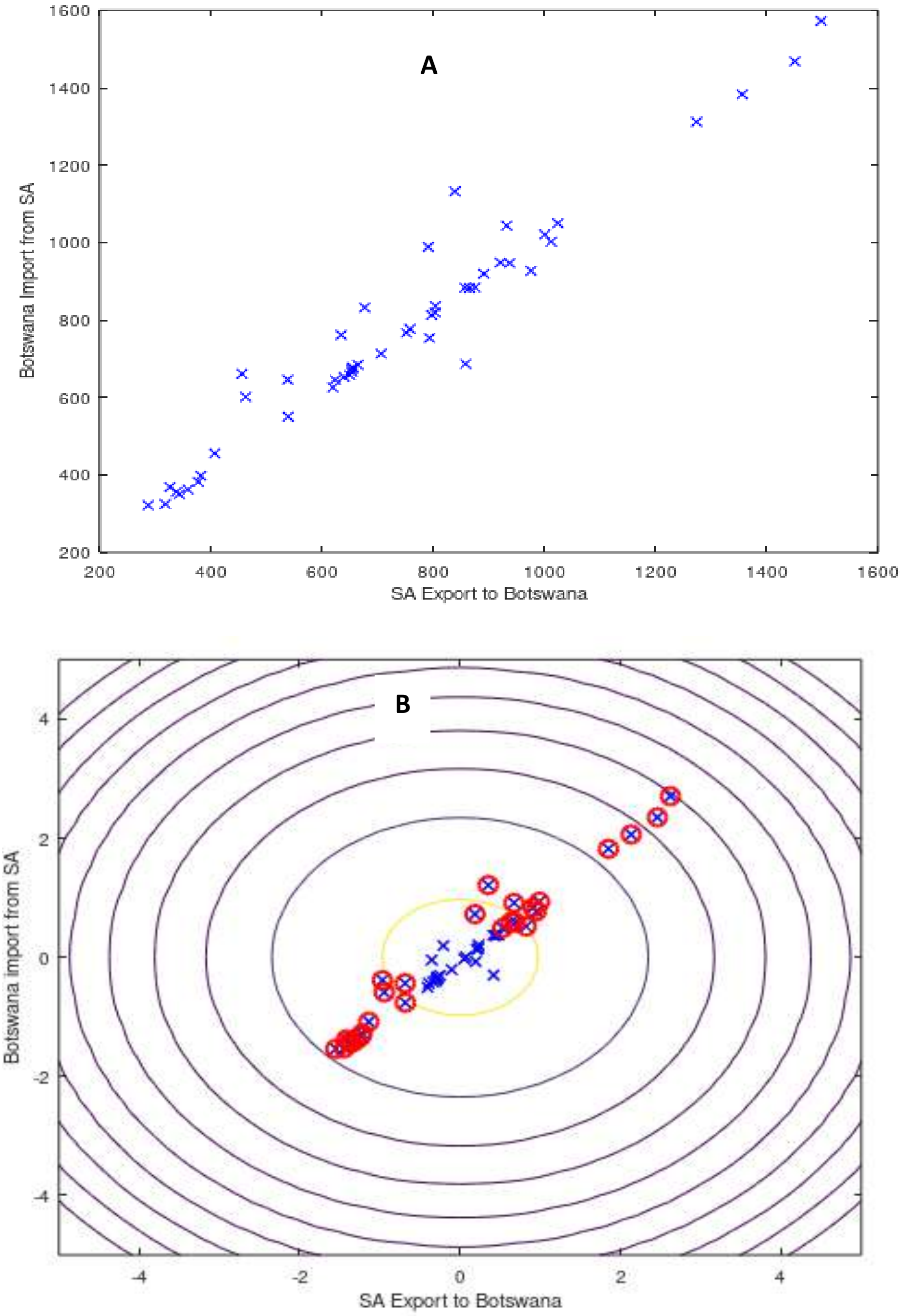


Figure 3: Simulated SA Export against Botswana Import (year 2000 to 2019)

The simulated randomized South African data represented in figure 3A shows fair linearity with the real data of Botswana imports from the year 2000 to 2019. In order to visualize the performance of the model, the values of the parameters of the model need to be considered. The model evaluates the mean, μ, of the netflow of randomized simulated South Africa's quarterly export to Botswana from 2000 to 2019 and the same corresponding real data of Botswana import from South Africa as 51.048 while the variance, σ^2, is 7916.4. Since it is assumed that the proximity of South Africa is considerably small, a trade reporting gap of 1.05% is considered for the analysis. The model evaluates the decision parameter, epsilon, on the cross-validation set as 0.3283345 for flagging whether or not the condition for illicit financial flow is positive or negative.

The model $P(X : \mu, \sigma^2) < \epsilon$ (0.3283345) flags the data points as illicit or non-illicit transactions as represented in Figure 3B with illicit transactions marked with red circles, while the respective values for mean and variance are substituted in the model equation as highlighted in Equation 1. A cluster of transaction data points shown in Figure 3B reflectively indicates a better outcome in terms of model performance in flagging illicit and non-illicit transactions.

4.3 Validation of GMAD Model using Simulation Data

To validate the model as effective in flagging whether a trade transaction is illicit or not, some data points are tested on the model to check if the model would flag it as illicit and the result of the analysis is compared with the result when Equation 4 is used and the condition for the illicit transaction is fulfilled.

$$\text{If } If\ (ABS\ (K_{i\,j} - \text{netflow for TBML}) > K_{i\,j}, \quad \begin{cases} 1 \text{ if true i.e flags anomaly} \\ 0 \text{ if false i.e flags non-anomaly} \end{cases}$$

The international trade flows, $K_{i\,j}$, is calculated as (FOB x 1.05) – FOB and this is a deciding factor for illicit (anomaly transactions) or normal (non-anomaly transactions). Equation 4 also defines the equation for evaluating the netflow for TBML.

Table 1: Model validation with the Test Set

FOB (SA)	Botswana	Netflow	$K_{i\,j}$	Illicit/Not	$P(X : \mu, \sigma^2)$	Model Result
1165	1350.50	185.5	67.525	1	0.295656	1
960.65	985.64	24.99	49.282	0	0.382841	0
1135.45	1189.53	54.08	59.477	0	0.411976	0
945.65	960.43	14.78	48.022	0	0.368435	0
1128.54	985.65	-142.89	56.427	1	0.088664	1
880.45	924.47	44.02	44.02	0	0.404082	0
756	782.46	26.46	37.80	0	0.384754	0
1125	1386.64	261.64	56.25	1	0.148440	1
867.56	1171.21	303.65	43.38	1	0.086833	1
965.65	989.79	24.14	48.28	0	0.381716	0

Table 1 represents the analysis of the simulated data of the South African export and corresponding Botswana imports from South Africa using the trade flow, $K_{i\,j}$, and the trade reporting gap conditions as compared with the model developed using GMAD on the same data. The model results parametrized by both the mean, μ, and variance, σ^2, is shown in the sixth (6th) column of Table 2 derived by applying equation 1 of the model.

$$P(X:\mu,\sigma^2)=\frac{1}{\sqrt{2\pi*15905.432}}\exp(-\frac{(x-79.637)^2}{2*(15905.4323})<\varepsilon(0.2958999) \quad (8)$$

Equation 8 represents the model equation for detecting illicit financial flow from simulated South Africa's export trade to Botswana and the corresponding Botswana import from South Africa, generated by the code that randomly selects numbers with a similar range of real data as presented by IMF's Trade Statistics database. The result is computed by applying Equation 8 for data in column six (6) and the outcome of the decision is shown in column seven (7) of Table 2. Similarly, the result when the same code used for analysing the real data is run on the simulated data is presented in Figure 4.

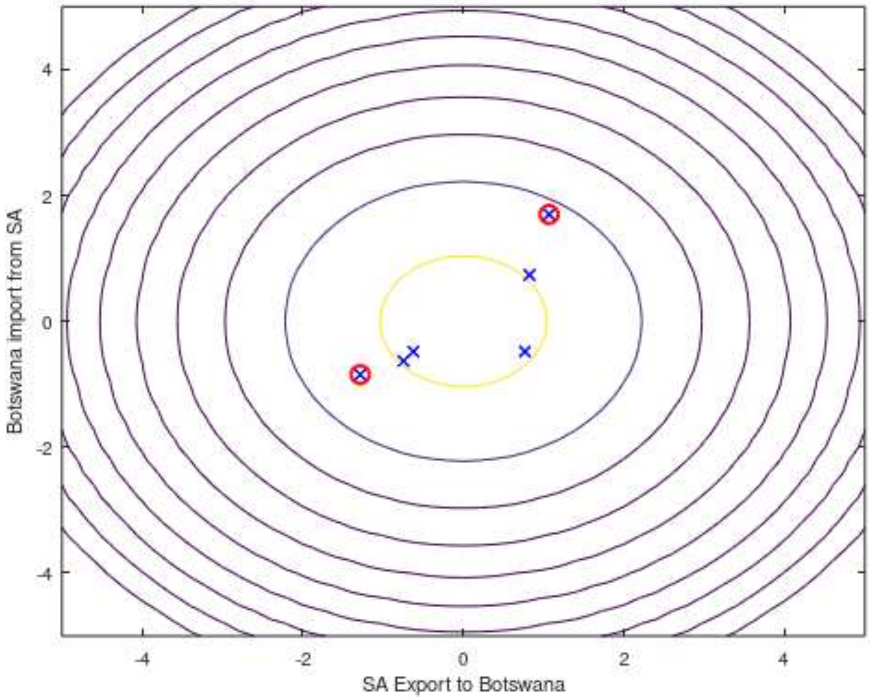


Figure 4: Detecting IFFs using GMAD model on Simulated Data

The result of the GMAD model on the simulated data shows that the model performs well in detecting whether or not a trade transaction is illicit or not, as the netfow of the trade transaction is taken as a parameter, x, in Equation 8, and the decision is reached based on the evaluation. The findings of the paper further confirm that machine learning techniques are able to predict illicit financial flows similar to the findings in literature (Wohl & Kennedy, 2018; Wong, 2015; Wu, 1997).

Practically, to apply this model for flagging if a trade transaction is illicit or not, the netflow is derived from the import and corresponding export as reported by the exporting country. This netflow is taken as variable 'x' of the model and substituted in Equation 8 in the case of trade transactions between South Africa and Botswana or similarly as the case may be. A positive result indicates that other factors such as tariff (for that year or period considered), corruption index, change in policies etc. may be instrumental in the result. If all factors are constant, it, therefore, affirms that the trade transaction is illicit.

5. Conclusion

Detecting illicit financial flow in trade transactions based on misreporting is of concern at the world economic forum and applying the machine learning method could present an option or alternative way of identifying IFFs. This paper adopts the GMAD model for the analysis of trade transactions between partner countries such as the US, Germany, China, South Africa, and Botswana, using real trade transaction data from IMF's Trade Statistics database. The result shows that trade data reflects some variance in transactions which is flagged as illicit by the model. Further research work may be done to investigate if there exist some factors that contribute to irregularities in trade transactions, to ascertain the effectiveness of the GMAD model in detecting IFFs.

The GMAD model shows a good result on the simulated data with the capability of detecting netflows that were illicit or not. The model may therefore be effective in detecting illicit financial flow in trade transactions if other factors such as tariffs, accounting and auditing standards, corruption, etc. that influence the trade data other than mispricing are held constant. Further deep machine learning methods could be used to detect and analyse trends in trade transactions to detect IFFs considering changing indicators that affect the performance of the model.

The largest gap in information exchange relates to trade misinvoicing. Therefore, we recommend the automatic exchange of information by trade and customs which most likely would address the anomalies in data misspecification.

The limitations of the study are that it is based on commodities recorded by both countries and focuses on illicit flows through values and quantities. However, there might be commodities that are reported by exporting countries and not recorded by importing countries. Evasion might not be the reason for the unrecorded data, therefore, overstating the evasion problem.

Future research may incorporate artificial intelligence and machine learning models using big data analytics and neural networks.

References

Africa, U. N. E. C. f. (2015). Illicit financial flow: report of the High Level Panel on illicit financial flows from Africa.

Baker, R. W. (2005). *Capitalism's Achilles heel: Dirty money and how to renew the free-market system*: John Wiley & Sons.

Bouet, A., & Roy, D. (2012). Trade protection and tax evasion: Evidence from Kenya, Mauritius, and Nigeria. *The Journal of International Trade & Economic Development, 21*(2), 287-320.

Cobham, A., & Janský, P. (2018). Global distribution of revenue loss from corporate tax avoidance: re-estimation and country results. *Journal of International Development, 30*(2), 206-232.

Davies, R. B., Martin, J., Parenti, M., & Toubal, F. (2018). Knocking on tax haven's door: Multinational firms and transfer pricing. *Review of Economics and Statistics, 100*(1), 120-134.

Doumbia, D., & Lauridsen, M. L. (2019). Closing the SDG Financing Gap.

Fisman, R., & Wei, S.-J. (2004). Tax rates and tax evasion: evidence from "missing imports" in China. *Journal of political Economy, 112*(2), 471-496.

Forstater, M. (2015). Can Stopping 'Tax Dodging'by Multinational Enterprises Close the Gap in Development Finance. *CGD Policy Paper, 69*, 32.

Forstater, M. (2017). Illicit financial flows and trade misinvoicing: Time to reassess. *Center for Global Development*.

Javorcik, B. S., & Narciso, G. (2008). Differentiated products and evasion of import tariffs. *Journal of International Economics, 76*(2), 208-222.

Javorcik, B. S., & Narciso, G. (2017). WTO accession and tariff evasion. *Journal of Development Economics, 125*, 59-71.

Kar, D., Cartwright-Smith, D., & Hollingshead, A. (2010). The Absorption of Illicit Financial Flows from Developing Countries: 2002-2006. *Available at SSRN 2335028*.

Kar, D., & Freitas, S. (2012). *Illicit financial flows from Developing Countries: 2001-2010*: Global Financial Integrity Washington, DC.

Kar, D., & LeBlanc, B. (2013). *Illicit financial flows from developing countries: 2002-2011*: Global Financial Integrity Washington, DC.

Kellenberg, D., & Levinson, A. (2019). Misreporting trade: Tariff evasion, corruption, and auditing standards. *Review of International Economics, 27*(1), 106-129.

Mishra, P., Subramanian, A., & Topalova, P. (2008). Tariffs, enforcement, and customs evasion: Evidence from India. *Journal of public Economics, 92*(10-11), 1907-1925.

Ndikumana, L. (2014). Savings, capital flight, and African development.

Nitsch, V. (2016). Trillion dollar estimate: Illicit financial flows from developing countries.

Rijkers, B., Baghdadi, L., & Raballand, G. (2017). Political connections and tariff evasion evidence from Tunisia. *The World Bank Economic Review, 31*(2), 459-482.

Rotunno, L., & Vézina, P. L. (2012). Chinese networks and tariff evasion. *The World Economy, 35*(12), 1772-1794.

Sequeira, S. (2016). Corruption, trade costs, and gains from tariff liberalization: Evidence from Southern Africa. *American Economic Review, 106*(10), 3029-3063.

Vézina, P.-L. (2015). Illegal trade in natural resources: Evidence from missing exports. *International Economics, 142*, 152-160.

Wang, C., & Han, D. (2019). Credit card fraud forecasting model based on clustering analysis and integrated support vector machine. *Cluster Computing, 22*(6), 13861-13866.

Wier, L. (2020). Tax-motivated transfer mispricing in South Africa: Direct evidence using transaction data. *Journal of Public Economics, 184*, 104153.

Wohl, I., & Kennedy, J. (2018). Neural network analysis of international trade. *US International Trade Commission: Washington, DC, USA*.

Wong, B. (2015). Do inflation expectations propagate the inflationary impact of real oil price shocks?: Evidence from the michigan survey. *Journal of Money, Credit and Banking, 47*(8), 1673-1689.

Worku, T., Mendoza, J. P., & Wielhouwer, J. L. (2016). Tariff evasion in sub-Saharan Africa: the influence of corruption in importing and exporting countries. *International Tax and Public Finance, 23*(4), 741-761.

Wu, R. C. (1997). Neural network models: Foundations and applications to an audit decision problem. *Annals of Operations Research, 75*, 291-301.

Yang, D. (2008). Can enforcement backfire? Crime displacement in the context of customs reform in the Philippines. *The Review of Economics and Statistics, 90*(1), 1-14.